N48uJ
no.1033

University of the State of New York Bulletin

Entered as second-class matter August 2, 1913, at the Post Office at Albany, N. Y., under the act of August 24, 1912. Acceptance for mailing at special rate of postage provided for in section 1103, act of October 3, 1917, authorized July 19, 1918

Published Fortnightly

No. 1033 ALBANY. N. Y. December 1, 1933

MAY 29 1934

THE PROGNOSIS OF TEACHING ABILITY OF STUDENTS IN NEW YORK STATE NORMAL SCHOOLS

BY

WARREN W. COXE, *Director*

AND

ETHEL L. CORNELL, *Psychologist*

Educational Research Division

New York State Education Department

ALBANY

THE UNIVERSITY OF THE STATE OF NEW YORK PRESS

1934

F30r-Mr34-1000(11988)

Return this book on or before the **Latest Date** stamped below. A charge is made on all overdue books.
University of Illinois Library

Dec 1

L161—H41

OF NEW YORK

llor - New York
Chan-
- - Binghamton
- - Yonkers
.L. - New York
- - Jamaica
- - Rochester
- - New York
- - Troy
- - Buffalo
- - Ogdensburg
.D. - Syracuse
.D. - New York

Education
D., LL.D.

L.D.

tion
LL.D.

ation
D., LL.D.

ation
LL.D.

n Education

D.

n

D.Sc.

Directors of Divisions

Archives and History, ALEXANDER C. FLICK M.A., Litt. D., Ph.D., LL.D.
Attendance and Child Accounting, CHARLES L. MOSHER Ph.M.
Educational Research, WARREN W. COXE B.S., Ph.D.
Examinations and Inspections, AVERY W. SKINNER B.A., Pd.D.
Health and Physical Education,
Law, CHARLES A. BRIND JR B.A., LL.B.
Library Extension, FRANK L. TOLMAN Ph.B., Pd.D.
Motion Picture, IRWIN ESMOND Ph.B., LL.B.
Professional Licensure, CHARLES B. HEISLER B.A.
Rehabilitation, RILEY M. LITTLE B.S., B.D.
Rural Education, RAY P. SNYDER
School Buildings and Grounds, JOSEPH H. HIXSON M.A.
Visual Instruction, ALFRED W. ABRAMS Ph.B.

C
N48uJ
no.1033
29 My 34 Hensel

FOREWORD

The real efficiency of a professional school for the education of teachers is based not upon buildings, equipment and sites but upon the abilities, interests and previous preparation of its students, upon its program of professional preparation and upon the professional attitude of staff members who are responsible for the instructional efficiency of the institution. The state normal schools and the State Teachers College at Buffalo offer three-year programs for the preparation of elementary school teachers which are based upon high school graduation and which require a quality of student performance worthy of the best efforts of the students admitted.

Until two years ago enrolments in these institutions were not restricted. Admission was based upon high school graduation or its equivalent. For the past two years enrolments have been restricted by the Commissioner of Education in accordance with each school's facilities for efficient instruction. The limitation has necessitated adoption of other criterions in addition to the completion of an approved high school course. Each applicant has, therefore, been required to have a high school Regents scholarship average of at least 72 per cent, a satisfactory rating in a general intelligence test, satisfactory ratings in certain specified subject matter tests, good health and good moral character.

This study of the prognosis of teaching ability of students in New York State normal schools interprets the following data concerning a group of entering students: (1) the results of certain entrance examinations; (2) the average of marks received at the close of certain semesters; and (3) the ratings of their teaching success at the close of the first and second years of public school teaching. The investigation shows the relationship between the criterions used in the selection of students, normal school success and later teaching success. It is the first study to follow a group of students through the training period and on into teaching, and the conclusions and recommendations are worthy of consideration and discussion by normal school and teachers college faculties. The authors raise some challenging questions relative to the criterions now used by the State in admitting candidates to professional study and with regard to present curriculum policies.

The State Education Department recognizes its indebtedness to Doctors Coxe and Cornell for this very valuable contribution to professional literature, and to the normal school faculty members and

THE UNIVERSITY OF THE STATE OF NEW YORK

Regents of the University
With years when terms expire

1944 James Byrne B.A., LL.B., LL.D., *Chancellor* - New York
1943 Thomas J. Mangan M.A., LL.D., *Vice Chancellor* - - - - - - - - - - - - - - - - Binghamton
1945 William J. Wallin M.A., LL.D. - - - - - Yonkers
1935 William Bondy M.A., LL.B., Ph.D., D.C.L. - New York
1941 Robert W. Higbie M.A., LL.D. - - - - - Jamaica
1938 Roland B. Woodward M.A., LL.D. - - - - Rochester
1937 Mrs Herbert Lee Pratt L.H.D. - - - - - New York
1939 Wm Leland Thompson B.A., LL.D. - - - - Troy
1936 John Lord O'Brian B.A., LL.B., LL.D. - - - Buffalo
1940 Grant C. Madill M.D., LL.D. - - - - - - Ogdensburg
1942 George Hopkins Bond Ph.M., LL.B., LL.D. - Syracuse
1946 Owen D. Young B.A., LL.B., D.C.S., LL.D. - New York

President of the University and Commissioner of Education
Frank P. Graves Ph.D., Litt.D., L.H.D., LL.D.

Deputy Commissioner and Counsel
Ernest E. Cole LL.B., Pd.D., LL.D.

Assistant Commissioner for Higher Education
Harlan H. Horner M.A., Pd.D., LL.D.

Assistant Commissioner for Secondary Education
George M. Wiley M.A., Pd.D., L.H.D., LL.D.

Assistant Commissioner for Elementary Education
J. Cayce Morrison M.A., Ph.D., LL.D.

Assistant Commissioner for Vocational and Extension Education
Lewis A. Wilson D.Sc.

Assistant Commissioner for Finance
Alfred D. Simpson M.A., Ph.D.

Assistant Commissioner for Administration
Lloyd L. Cheney B.A., Pd.D.

Director of State Library
James I. Wyer M.L.S., Pd.D.

Director of Science and State Museum
Charles C. Adams M.S., Ph.D., D.Sc.

Directors of Divisions

Archives and History, Alexander C. Flick M.A., Litt. D., Ph.D., LL.D.
Attendance and Child Accounting, Charles L. Mosher Ph.M.
Educational Research, Warren W. Coxe B.S., Ph.D.
Examinations and Inspections, Avery W. Skinner B.A., Pd.D.
Health and Physical Education,
Law, Charles A. Brind jr B.A., LL.B.
Library Extension, Frank L. Tolman Ph.B., Pd.D.
Motion Picture, Irwin Esmond Ph.B., LL.B.
Professional Licensure, Charles B. Heisler B.A.
Rehabilitation, Riley M. Little B.S., B.D.
Rural Education, Ray P. Snyder
School Buildings and Grounds, Joseph H. Hixson M.A.
Visual Instruction, Alfred W. Abrams Ph.B.

C
N48uJ
no. 1033
29 My 34 Hensel

FOREWORD

The real efficiency of a professional school for the education of teachers is based not upon buildings, equipment and sites but upon the abilities, interests and previous preparation of its students, upon its program of professional preparation and upon the professional attitude of staff members who are responsible for the instructional efficiency of the institution. The state normal schools and the State Teachers College at Buffalo offer three-year programs for the preparation of elementary school teachers which are based upon high school graduation and which require a quality of student performance worthy of the best efforts of the students admitted.

Until two years ago enrolments in these institutions were not restricted. Admission was based upon high school graduation or its equivalent. For the past two years enrolments have been restricted by the Commissioner of Education in accordance with each school's facilities for efficient instruction. The limitation has necessitated adoption of other criterions in addition to the completion of an approved high school course. Each applicant has, therefore, been required to have a high school Regents scholarship average of at least 72 per cent, a satisfactory rating in a general intelligence test, satisfactory ratings in certain specified subject matter tests, good health and good moral character.

This study of the prognosis of teaching ability of students in New York State normal schools interprets the following data concerning a group of entering students: (1) the results of certain entrance examinations; (2) the average of marks received at the close of certain semesters; and (3) the ratings of their teaching success at the close of the first and second years of public school teaching. The investigation shows the relationship between the criterions used in the selection of students, normal school success and later teaching success. It is the first study to follow a group of students through the training period and on into teaching, and the conclusions and recommendations are worthy of consideration and discussion by normal school and teachers college faculties. The authors raise some challenging questions relative to the criterions now used by the State in admitting candidates to professional study and with regard to present curriculum policies.

The State Education Department recognizes its indebtedness to Doctors Coxe and Cornell for this very valuable contribution to professional literature, and to the normal school faculty members and

public school supervisors, who have given constructive assistance in the collection of the data used throughout the study. The State Department turns again to normal school and teachers college faculties for assistance in the development of a program of admissions which will give the normal school a type of student most likely after three years of professional study to render the State satisfactory service in its public schools.

HERMANN COOPER

Assistant Commissioner for Teacher Education and Certification

CONTENTS

PAGE

Introduction 7

I Results of the entrance tests 10

Range of test results in each school 10

Variations among schools as to students in the highest and lowest tenths 17

Tendencies for students' ratings to be consistently low or high in various measures 19

Low test scores as a basis for selective admissions 21

II How the entrance tests were related to normal school work 23

Correlations between entrance measures and achievement in normal school 23

Analysis of normal school marks 29

III How the entrance tests were related to rating of teaching success 38

Prognostic value of entrance tests as indicated by correlations with the Ten-Point Rating 39

Comparison of various ratings of teachers 39

Prognostic value of entrance tests as indicated by correlations with Almy-Sorenson and Torgerson ratings 41

Prognostic value of combined low scores on entrance tests 42

IV What the criterions measure 43

Validity of criterions as judged by differences between good and poor teachers selected by these criterions 43

Differences between best and poorest teachers 46

Differences between good and poor teachers 52

V Conclusions 57

University of the State of New York Bulletin

Entered as second-class matter August 2, 1913, at the Post Office at Albany, N. Y., under the act of August 24, 1912. Acceptance for mailing at special rate of postage provided for in section 1103, act of October 3, 1917, authorized July 19, 1918

Published Fortnightly

No. 1033 ALBANY, N. Y. December 1, 1933

THE PROGNOSIS OF TEACHING ABILITY OF STUDENTS IN NEW YORK STATE NORMAL SCHOOLS

INTRODUCTION

For a number of years the normal schools of the State have been considering the problem of how to select students for teacher training in such a way as to reduce the number of normal school graduates who fail to make a success of teaching. In 1927 a battery of tests was given to all entering students in the ten normal schools of the State (including the State College for Teachers at Buffalo, which still had a three-year training course at that time) in order to determine the prognostic value of entrance tests in such a selective admissions program. This study was undertaken with the purpose of following the students through normal school and one or two years of teaching so that there should be actual teaching experience for a criterion and not only practice teaching or normal school work.

The entrance measures used were the following:

1 Terman Group Test of Mental Ability[1]

2 Tressler Minimum Essentials in English Test[2]

3 Whipple High School and College Reading Test[2]

4 Teachers' Interest Test — devised for this purpose

5 Prognosis Test — devised for this purpose and since published[1]

6 High school marks — letter grades transmuted into numerical scores

7 New York Rating Scale for School Habits[1] (which was filled out for each candidate by the principal or teacher of the high school from which the student came)

[1] Published by World Book Company, Yonkers, N. Y.
[2] Published by Public School Publishing Company, Bloomington, Ill.

In addition to the entrance tests the following data were gathered for as many individuals as possible:

a First semester marks in "content" subjects

b First semester marks in "professional" subjects

c Comprehensive achievement test of first year's work (objective test devised for the purpose) and referred to as "Second Semester Achievement Test"

d Sixth semester marks in "content" subjects

e Sixth semester marks in "professional" subjects

f Marks in practice teaching (fifth or sixth semester)

g Rating of teaching ability at end of first year of teaching on *Ten-Point Rating Scale* (devised for the purpose) by local supervisor, principal or superintendent

h Report from teachers themselves as to difficulties experienced in first year

i Report from local supervisors as to difficulties experienced by teachers in first year

j Rating of teaching ability at end of second year of teaching on *Ten-Point Rating Scale* by local supervisor

k Rating of teaching ability (for selected teachers) at end of first year of teaching, by two expert raters who visited class, on *Torgerson Diagnostic Rating Scale of Teaching Ability* (a rating of teaching technic)

l Rating of same group by same raters on *Almy–Sorenson Rating Scale for Teachers* (a rating of teaching personality)

m Rating of same group by local supervisors on *Almy–Sorenson Rating Scale for Teachers*

n A *general rating* of same group by local supervisors (opinion of supervisor as to whether teacher was good, average, or poor)

The original group examined was composed of about 1850 entering students. The number of these who received sixth semester marks three years later had dwindled to a few more than 950. The number who were located in teaching positions in their first year (1930–31) was a little less than 800. The number for whom we have ten-point ratings in the second year is about 400. The number who were personally visited by two expert raters is a little over 100.

We may say at the outset that our effort at prognosis was to some extent defeated by the lack of agreement among judges as to which teachers concerned in this study were most successful. We may say

also that we found very few teachers who were regarded as definitely unsuccessful. It appears that some selective factor had already been at work before the teachers actually got into the field. To what extent the elimination of half the number of entering students before securing jobs was elimination of those who would probably be least successful we do not know. Much of it, perhaps most, was not elimination from teaching, but was due to the practice common at the time of leaving normal school after two years to teach on a limited certificate. Unfortunately we have no data to indicate whether the students who were eliminated for this reason were selected as to their scholarship or as to their teaching ability. We may assume, however, that such elimination as occurs is more wasteful not only in terms of public funds but also in terms of the happiness and usefulness to society of the individuals eliminated than would be a method of selecting students at or before entrance to normal school, if the latter could be done.

While our study has been disappointing in the extent to which we can make definite recommendations as to bases for admission of students, as indeed have been all other studies of a similar nature, we feel that it is worth while to report the study in as brief a form as can be done intelligibly, for the light it throws not only on the problem of prognosis but on the whole vexed problem of the measurement of teaching success.

The main part of the report is divided into four sections. The first section gives a general summary of the test results for the entering students. The second section shows how the entrance tests were related to achievement in the normal school. In the third section we discuss the relation of the tests to the various measures of teaching success and show to what extent the tests are prognostic of success as measured in various ways and by a composite criterion. In the fourth section the problem of the validity of the criterions is approached by a study of the differences between the best and poorest teachers selected on the basis of these criterions.

RESULTS OF THE ENTRANCE TESTS

Range of test results in each school. A general picture of the results of the entrance tests is given by figures 1–7, which show the range of score from the tenth to the ninetieth percentile in each school in the seven entrance measures used, namely: Terman Intelligence, Tressler English, Whipple Reading, Teachers' Interest, Prognosis, High School Marks and Habit Rating Scale. When standards for comparison are available, they are indicated at the right of the graph. Each school is distinguished by a code number which will be used throughout the study.

It may be noted that on the Terman Intelligence Test almost 70 per cent of the normal school entrants for the State as a whole exceeded the median for the last year of high school. In only two schools was this point (the thirtieth percentile) lower than the high school median.[1] The tenth percentile for every school is over ten points of score higher than the tenth percentile for high school seniors, although the ninetieth percentile for high school seniors almost reaches the ninetieth percentile for normal school entrants for the State as a whole, and exceeds that of one normal school. That is, the normal school entrants are definitely selected as to intelligence compared with students in the twelfth grade. There are greater differences in the scores attained by the lowest 10 per cent and the lowest 30 per cent of the different normal schools than there are in the medians or higher percentiles of the schools. Evidently some schools have to meet the needs of a wider range of student ability than others.

In the Tressler English Test there are greater differences between schools in the median and higher percentiles than in the lower. In this test the normal schools are very far below the standards for

[1] The use of percentile comparisons may need a word of explanation. In each test we have arranged the scores in order, and divided the number of students into ten groups, each group containing 10 per cent of all the students taking the test. The dividing lines between groups are therefore the scores attained by the tenth, twentieth, thirtieth ... to ninetieth student out of every hundred, when scores are arranged from lowest to highest. The lowest tenth or lowest 10 per cent of the students are then at or below the tenth percentile point in score, the lowest 30 per cent or lowest three-tenths are at or below the thirtieth percentile, the highest 10 per cent are above the ninetieth percentile, the highest 30 per cent are above the seventieth percentile, etc. Schools may then be compared either by noting the difference in score between the score made by the lowest 10 per cent of the school's enrolment and the score made by the lowest 10 per cent of all students, or by noting what per cent of the school's enrolment reaches the tenth percentile score of the whole group. If 30 per cent of one school are below the tenth percentile of the whole group, obviously that school has a larger number of poor students than the average; if only 5 per cent of one school are below the tenth percentile, that school has a smaller number of poor students.

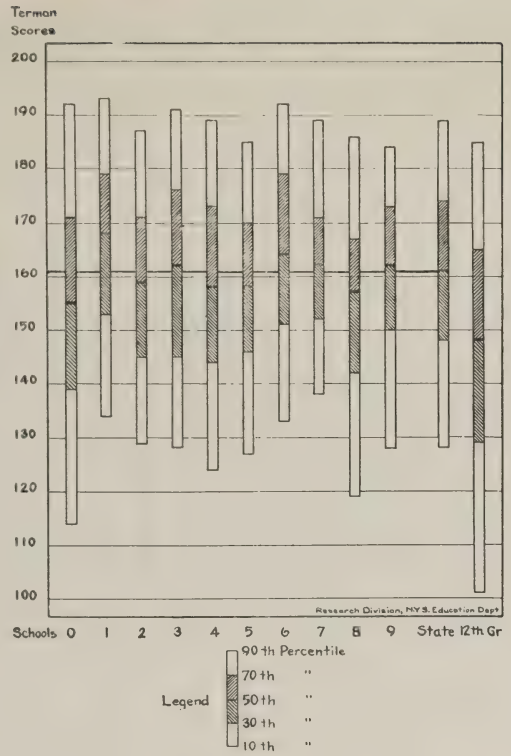
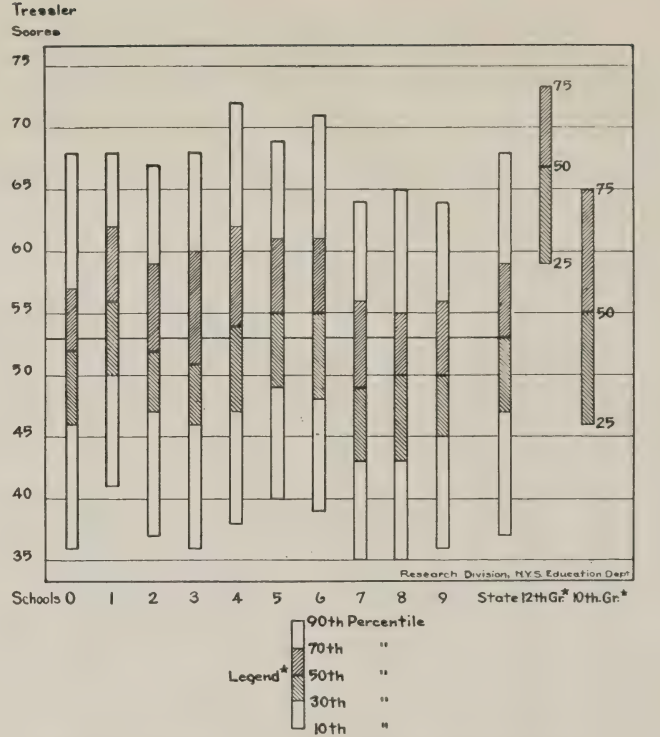


Figure 1 Terman Group Test of Mental Ability. Range for each normal school from tenth to ninetieth percentile.



*Note: High school percentiles are 75th, 50th and 25th instead of 90th, 70th, etc.

Figure 2 Tressler Minimum Essentials in English Test. Range for each normal school from tenth to ninetieth percentile.

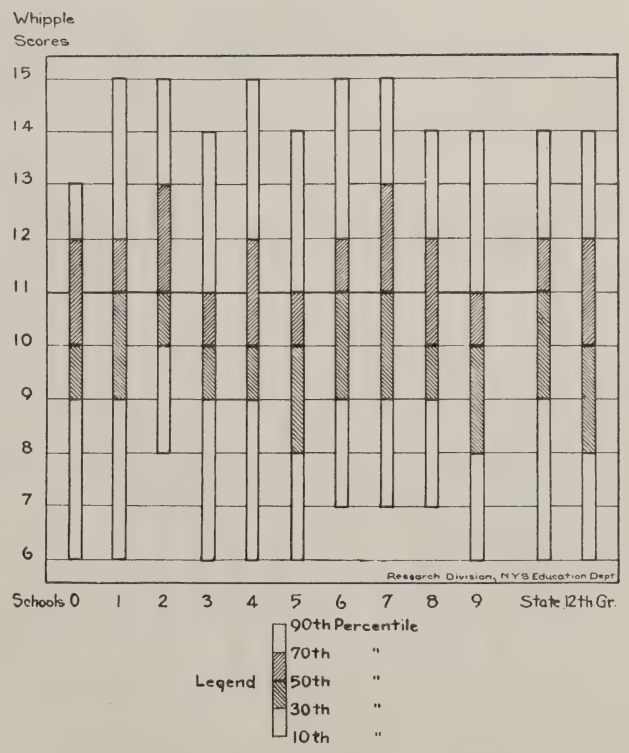


Figure 3 Whipple High School and College Reading Test. Range for each normal school from tenth to ninetieth percentile.

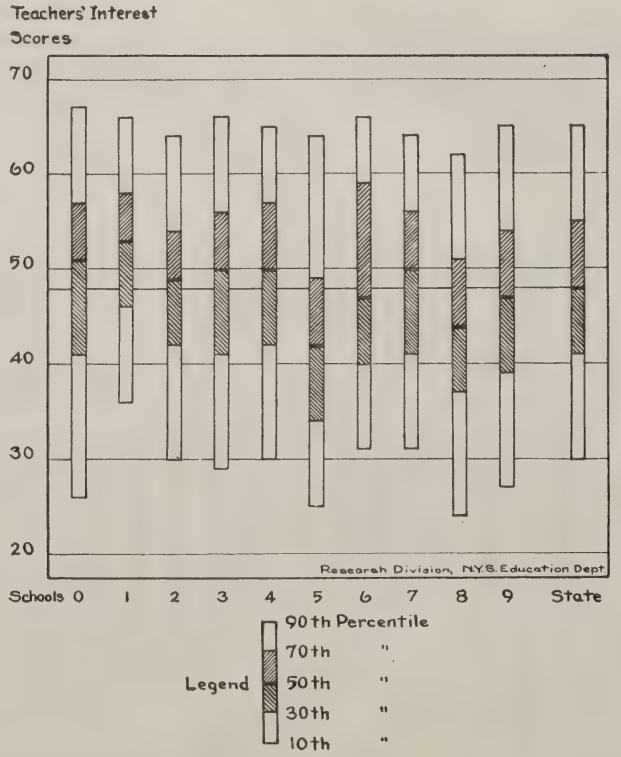


Figure 4 Teachers' Interest Test. Range for each school from tenth to ninetieth percentile.

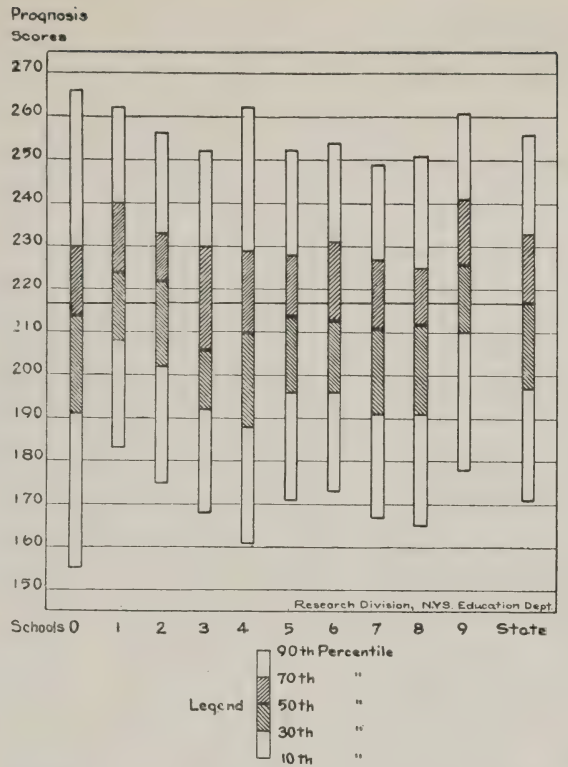


Figure 5 Prognosis Test for Teachers. Range for each normal school from tenth to ninetieth percentile.

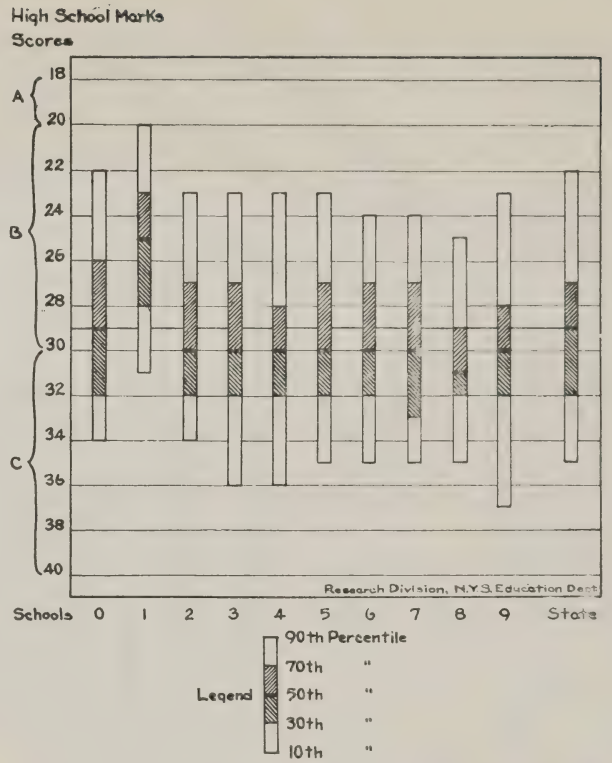


Figure 6 High School Marks. Range for each normal school from tenth to ninetieth percentile.

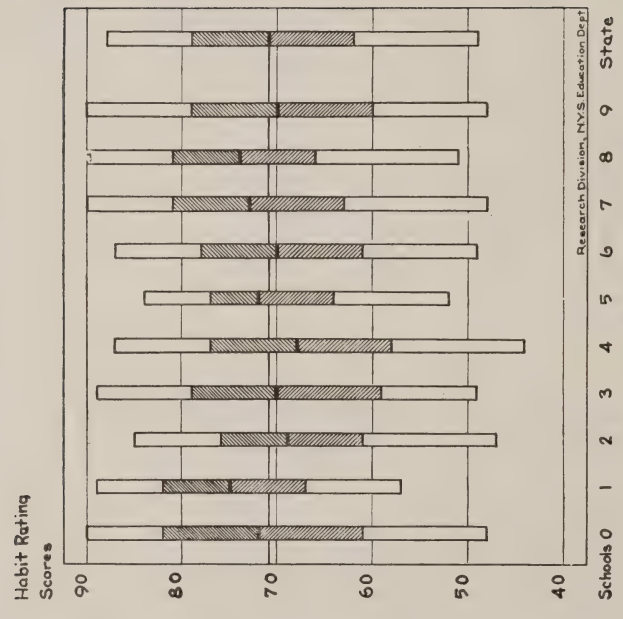


Figure 7 New York Rating Scale for School Habits. Range for each normal school from tenth to ninetieth percentile.

twelfth grade: the ninetieth percentile for the normal schools of the State as a whole is barely above the median for the twelfth grade, and the median for the normal schools is well below the twenty-fifth percentile for the twelfth grade, and even somewhat below the median for the tenth grade. Evidently, these normal school students tended to be among the poorer high school students in English, as measured by this test.

In the Whipple Reading Test there is a great deal of variation among the different schools in the range of scores. For the State as a whole, the range from tenth to ninetieth percentiles is practically the same as the high school range but the median is a little higher than the high school median.

The Teachers' Interest Test and the Prognosis Test were devised for the purposes of this study and have therefore no external criterions for comparison. It is rather curious that the schools which are below the average of the State, both at the median point and at the seventieth percentile, in one of these tests tend to be above that point in the other test. Whether there is really any compensating factor at work or whether this is a chance result can not be stated.

High school marks were turned into numerical values from one to five and averaged (the resulting decimal being omitted), so that values of 10–19 indicate A's; 20–29, B's; 30–39, C's; etc. It is evident that very few A averages are found among normal school entrants except in one school and that the largest number of marks is B.

The Habit Rating Scale was filled out by the high school from which the student came, since the normal schools were not sufficiently acquainted with the students to rate them at the time of entrance. Considering the diversity of persons concerned in the ratings, it is surprising to find so little difference between the schools. The differences among either the highest or the lowest tenth, however, are much greater than the differences in the middle of the range. Few pupils in any school have ratings which would indicate only average possession of the nine traits rated. The ninetieth percentile in four schools is at the highest score possible and the ninetieth percentile for all schools together is very close to the highest possible score. Unfortunately we do not know how twelfth grade students in general would be rated on this scale, but at least we can say that normal school entrants are rated very high.

The number of students involved in these entrance measures is shown in table 1. The measures give a fairly complete sampling of students except in the case of high school marks for School 8, where

TABLE 1

Number of students taking each entrance test

ENTRANCE MEASURES	NORMAL SCHOOLS										TOTAL
	0	1	2	3	4	5	6	7	8	9	
Terman	71	251	194	191	210	202	178	157	141	213	1 808
Tressler	71	251	242	191	210	202	177	161	141	213	1 859
Whipple	71	248	249	191	210	202	178	161	141	212	1 863
Teachers' Interest	71	261	232	191	170	201	177	158	141	213	1 825
Prognosis	71	256	232	190	167	201	177	159	141	213	1 807
High School Marks	69	246	237	187	155	189	171	150	44	211	1 659
Habit Rating	67	208	218	178	156	188	158	143	128	191	1 635

the data are given for only about a third of the students. The largest number of students involved in any of the measurements was 1863 (Whipple Reading) and the smallest number was 1635 (Habit Rating Scale).

Variations among schools as to students in the highest and lowest tenths. The graphs presented show that the normal schools differ in the character of their best and poorest students as much as or more than in that of their average students. We can present a little more clearly the facts regarding the most extreme students at each school by showing the percentage of the students of each school which exceeded the ninetieth percentile of all students and the percentage which fell below the tenth percentile of all students. The per cent in different schools falling in the lowest tenth of all students on the Terman Test varies from 4.2 per cent in School 7 to 18.7 in School 0 (table 2). That is, School 0 has almost four times as large a proportion of the poorest students (poorest in terms of an intelligence test) as School 7. In high school marks, School 1 has only 1.3 per cent of the lowest tenth, while School 3 has 14.9 per cent. It will be noticed by comparing the percentages for one school in different entrance measures (vertically instead of horizontally in the table) that no school has either more or less than 10 per cent of its students in the lowest tenth on every measure. School 0 comes nearest to having more than its share of low students in all tests. More than 10 per cent of its students are in the lowest tenth in all measures except high school marks. School 1 seems to be most consistent in having less than 10 per cent in the lowest tenth, Whipple Reading being the only measure in which it has more than 10 per cent. School 6 also has less than 10 per cent in all measures except high school marks.

In regard to the proportion of students who exceed the ninetieth percentile (that is, are in the highest tenth) of all students, we find the greatest variation in high school marks. The percentage of each school which falls in the highest tenth varies from 0.5 per cent of School 8 to 25.5 per cent of School 1. At this end of the range we find that School 8 is most likely to have less than 10 per cent of the best pupils on all tests, and School 1 to have more than 10 per cent. Schools with a high percentage of students in the lowest tenth are not always those with the lowest percentage in the highest tenth. School 0, which had many poor students, also tends to have many of the best. That is, the normal schools differ in the *range* of ability of their students, some schools having a more heterogeneous student body than others.

TABLE 2

Per cent of each normal school falling in lowest and highest tenths on each entrance test

ENTRANCE MEASURES	IN LOWEST TENTH OF ALL STUDENTS									
	Schools									
	0	1	2	3	4	5	6	7	8	9
Terman	18.7	6.6	9.1	10.7	8.2	10.8	4.8	4.2	16.7	9.9
Tressler	12.6	4.6	10.1	10.3	8.8	6.9	8.0	15.1	14.3	13.7
Whipple	15.5	11.7	4.4	10.6	9.6	15.0	6.4	5.6	8.1	13.1
Teachers' Interest	13.6	2.7	9.1	11.1	9.0	18.1	7.6	7.3	18.5	12.7
Prognosis	16.9	6.2	7.4	12.0	16.1	9.9	8.8	11.9	13.5	5.7
High School Marks	7.5	1.3	8.9	14.9	13.9	11.0	12.4	12.5	8.2*	11.0
Habit Rating	10.9	5.9	12.5	10.0	15.6	6.5	9.3	10.8	8.4	11.7
	IN HIGHEST TENTH OF ALL STUDENTS									
Terman	12.7	13.7	7.7	11.5	13.5	5.5	14.0	9.4	7.4	5.2
Tressler	9.4	10.7	8.9	9.8	17.5	10.7	13.1	4.9	6.4	6.0
Whipple	5.4	10.5	12.4	7.2	12.6	8.8	11.0	14.3	8.1	5.2
Teachers' Interest	12.7	11.8	8.8	11.2	10.3	12.0	10.9	9.1	6.5	9.7
Prognosis	13.2	13.0	9.8	7.6	12.6	7.2	8.9	6.4	7.0	14.1
High School Marks	10.7	25.5	7.6	7.1	6.6	8.4	6.0	7.0	0.5*	8.3
Habit Rating	14.8	12.5	5.5	9.3	8.2	3.7	8.6	14.6	12.6	13.6
Approximate number	71	251	194	191	210	202	178	157	141	213

* Only about one-third of the students had high school marks reported.

Tendencies for students' ratings to be consistently low or high in various measures. The most usual method of reporting test relationships is by presenting the intertest correlations. These were not worked out for the whole student body but are available for a selected group of 89 cases who were followed through graduation and a year of teaching. These correlations, shown in table 3, are presented merely as a means of general orientation to the degree of

TABLE 3

Intertest correlations

	TERMAN	TRESSLER	WHIPPLE	TEACHERS' INTEREST	PROGNOSIS	HIGH SCHOOL MARKS
Terman	.....	.....	.....	.....	.....	.....
Tressler	.55	.....	.....	.....	.....	.....
Whipple	.52	.33	.....	.....	.....	.....
Teachers' Interest	.50	.30	.37	.....	.....	.....
Prognosis	.66	.52	.51	.42	.....	.....
High School Marks	.52	.53	.32	.41	.50	.....
Habit Rating	.23	.21	.08	.14	.33	.35

NOTE. These correlations are based on 89 cases who were graduated, obtained teaching positions, and were rated in their first year of teaching. They are obviously not a random sampling of entering students. They are the cases involved in a multiple correlation with teaching success. The correlations are in general somewhat higher than were obtained in a random sampling of cases used to analyze the Prognosis Test.

relationship between tests, which we shall take up from a somewhat different angle. No particular comment is needed on these correlations; they are of the usual order obtained by such testing programs. The tests correlate more highly with each other and with school marks than with Habit Rating marks.

For the purpose of considering what possibilities for selective admissions might be found in low test scores, which is our present purpose, it is more pertinent to inquire how many students who are low in one measure are also low in others, and how many are high. Table 4 shows the percentage of the lowest tenth of students on each entrance measure which fell in the lowest three-tenths (below the thirtieth percentile) on each other measure, and the percentage which fell in the highest three-tenths (above the seventieth percentile) on each other measure. The Terman Intelligence Test is the measure on which those who are in the lowest tenth are the most likely to fall below the thirtieth percentile in other measures and, conversely, least

TABLE 4

Consistency of entrance measures for students with low scores. Percentage of scores of lowest tenth in one test falling in lowest three-tenths and highest three-tenths on other tests

PER CENT FALLING IN LOWEST THREE-TENTHS ON	LOWEST TENTH OF TESTS						
	Terman	Tressler	Whipple	Teachers' Interest	Prognosis	High School Marks	Habit Rating
Terman	....	78.4	66.0	62.1	81.7	64.5	40.0
Tressler	70.8	....	59.6	54.9	64.2	59.9	43.8
Whipple	80.5	67.8	....	70.5	75.8	67.2	49.1
Teachers' Interest	61.9	53.6	53.0	....	62.2	48.8	41.2
Prognosis	87.9	68.0	66.7	54.6	....	62.7	42.0
High School Marks	55.3	49.7	40.0	44.3	53.8	....	39.8
Habit Rating	48.0	46.1	33.9	42.2	41.6	46.5	....
Average	67.4	60.6	53.2	54.8	63.2	58.3	42.7

PER CENT FALLING IN HIGHEST THREE-TENTHS ON	LOWEST TENTH OF TESTS						
	Terman	Tressler	Whipple	Teachers' Interest	Prognosis	High School Marks	Habit Rating
Terman	....	4.4	5.1	4.6	1.8	4.7	21.2
Tressler	0.6	....	12.4	10.8	5.3	3.9	17.3
Whipple	0.6	3.9	....	4.6	0.6	8.6	10.6
Teachers' Interest	7.5	11.6	14.2	....	7.1	16.8	21.8
Prognosis	0.6	8.8	7.3	8.7	....	10.3	21.8
High School Marks	2.5	8.5	14.6	14.1	10.0	....	14.3
Habit Rating	16.0	13.8	19.8	18.9	17.5	17.5	....
Average	4.6	8.5	12.2	10.3	7.1	10.3	17.8

likely to be found among the best three-tenths of students in any other measure. The Prognosis Test is the second most consistent measure for selecting poor students, and the Tressler English Test is third. But on the average, a third or more of the students in the lowest 10 per cent of any one of these three tests will be above the thirtieth percentile on each of the other measures. Less than 10 per cent of these poorest students, however, are to be found among the highest tenth of the students on other tests. The Habit Rating Scale is evidently the least related to the other measures. On the average, only a little over 40 per cent of the lowest tenth in Habit Rating are below the thirtieth percentile on other tests, while about 18 per cent of them are above the seventieth percentile.

Table 5 shows the same facts for each normal school separately for certain of the measures. The number of cases involved is small, however, and the differences between schools are not significant unless they are large. In general, differences found do not reach the criterion of statistical validity.

Low test scores as a basis for selective admissions. Since we find that some students who are among the lowest on any one of these measures are among the highest on some other measure, it would obviously be unfair to consider the possibility of eliminating candidates on the basis of any one measure alone. It is apparent from table 2 that one normal school does (or did at this time) eliminate candidates largely on the basis of high school marks, yet some students with the lowest marks (of whom there are only 1.3 per cent in that school) are found among the best students in other tests.

We therefore found for each school the percentage of the whole number of entering students who were in the lowest tenth on any test, on any two tests, on any three tests etc. Table 6 shows these percentages. As they are based on the whole number of students in each school, the differences between schools are much more likely to be significant than in table 5. We find differences among schools ranging from about one-quarter of all students in one school who are in the lowest tenth in at least one test to one-half of the students in other schools. If it should prove to be desirable to eliminate approximately 10 per cent of the present enrolment (as was regarded as probable at the beginning of this inquiry) it would apparently be possible to eliminate about this number of the poorest *students* by taking those who were in the lowest tenth on at least three of the measures. Whether these would necessarily prove to be the poorest *teachers* after they were graduated, we could not determine at this stage.

TABLE 5

Percentage of scores of lowest tenth in each normal school on Terman, Tressler and Prognosis tests falling in lowest three-tenths on other tests

SCHOOL	TERMAN — LOW TENTH				
	In lowest three-tenths on				
	Tressler	Whipple	Prognosis	High School Marks	Habit Rating
0	77	85	100	54	33
1	77	77	88	7	38
2	83	75	92	50	57
3	57	74*c*	82	78	42
4	59	76	90	56	54
5	52*a*	87	83	73	35
6	83	75	83	50	56
7	50*b*	100*c*	75	75	25
8	80*b*	80	100	60	80
9	92*a*	92	87	48	64

SCHOOL	TRESSLER — LOW TENTH				
	In lowest three-tenths on				
	Terman	Whipple	Prognosis	High School Marks	Habit Rating
0	88	75	100	72	43
1	60*d*	80	53*f*	..	33
2	75	58	57	41	46
3	79	67	75	57	52
4	86*e*	57	81*f*	43	38
5	88	76	65	64	38
6	79*d*	57	62	69	50
7	88	67	78	56	25
8	50*e*	50	75	100	75
9	79	81	67	53	58

NOTE. Indications of the reliability of the differences are given by the following illustrations:

a The difference between 92 per cent and 52 per cent (40) is more than three times the standard error of the difference.

$$\sigma d = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

b The difference between 50 per cent and 80 per cent (30) is just equal to the standard error of the difference.

c The difference between 74 per cent and 100 per cent (26) is less than the standard error of the difference.

d The difference between 79 per cent and 60 per cent (19) is just slightly larger than the standard error.

e The difference between 86 per cent and 50 per cent (36) is 1.4 times the standard error.

f The difference between 81 per cent and 53 per cent (28) is 1.8 times the standard error.

g The difference between 90 per cent and 60 per cent (30) is 1.8 times the standard error.

TABLE 5 (*concluded*)

Percentage of scores of lowest tenth in each normal school on Terman, Tressler and Prognosis tests falling in lowest three-tenths on other tests

SCHOOL	PROGNOSIS — LOW TENTH				
	In lowest three-tenths on				
	Terman	Tressler	Whipple	High School Marks	Habit Rating
0	100	62	69	46	58
1	68	53	68	6	33
2	83	83	56	53	41
3	72	72	72	54	41
4	89	74	82	83	55
5	90*g*	43	86	65	40
6	81	50	75	60	36
7	60*g*	55	82	55	50
8	75	50	88	38	38
9	92	92	91	50	33

SCHOOL	TEACHERS' INTEREST — LOW TENTH					
	In lowest three-tenths on					
	Terman	Tressler	Whipple	Prognosis	High School Marks	Habit Rating
0	80	70	80	80	50	30
1	58	75	58	58	..	28
2	52	52	52	43	48	36
3	62	58	58	46	27	57
4	81	56	88	81	64	79
5	66	40	81	53	44	34
6	53	53	53	53	42	53
7	67	83	83	67	83	33
8	86	57	72	86	57	43
9	49	61	79	42	42	36

HOW THE ENTRANCE TESTS WERE RELATED TO NORMAL SCHOOL WORK

Correlations between entrance measures and achievement in normal school. The first step in determining the relation of the entrance measures to the work of the student in normal school was to correlate the entrance measures with normal school marks. This has been done for the average of first semester marks in "content" subjects and in "professional" subjects, for the average of sixth

TABLE 6

Percentage of students taking tests in each normal school who were in lowest tenth on one or more tests

SCHOOL	1 OR MORE	2 OR MORE	3 OR MORE	4 OR MORE	5 OR MORE	6
0	43.6*a*	23.8	14.0	9.8	4.2	1.4
1	27.7*a*	9.8	3.9	1.2	.4	....
2	39.4	16.2	7.9	3.3	1.2	.4
3	48.2	24.0	10.4	3.6	.5	....
4	50.9	25.2*b*	13.2	7.8	2.4	....
5	43.1	25.6	13.3	4.0	2.0	....
6	37.8	15.6	6.8	2.3	.6	....
7	50.0	12.1*b*	5.5	2.2	1.1	....
8	29.2	23.4	12.8	6.4	....	....
9	46.0	20.7	11.8	6.6	1.9	....

NOTE. Illustrations of the degree of reliability of differences are given by the following:
a Difference between 43.6 per cent and 27.7 per cent is exactly four times its P.E.
b Difference between 25.2 per cent and 12.1 per cent is 4.4 times its P.E.

semester marks in these two groups and for the mark in practice teaching, which was sometimes given in fifth semester and sometimes in sixth. In order to have some objective measure of normal school work which would be comparable in all schools, the first year's work at the end of the second semester was measured by an objective test constructed for this purpose. The correlations for each school are shown in tables 7, 8 and 9 and the average of various groups of correlations is presented in table 10.

Perhaps the most striking point about these correlations is the extent to which they vary in different schools. Within a single school, correlations of normal school marks with each of the entrance measures tend to cluster within a fairly narrow range around their central tendency, but the correlations of school marks with any single entrance test in the different schools show no consistency whatever. An illustration of this may be seen in table 7, A, Schools 3 and 4. In School 3, correlations of first semester professional marks with the entrance tests are below .10 in four cases and between .15 and .20 in three, the average being .11. In School 4, six of the correlations are between .30 and .40 and one is .52, the average being .39. In School 3 the Terman Test correlates with marks to the extent of .04 and in School 4 to the extent of .40, and the Prognosis Test in School 3 has a correlation of .05 and in School 4 of .52. A similar trend is observable for these two schools in the first semester correlations with marks in content courses, although not so extreme, and also in sixth semester correlations with professional marks. In

TABLE 7

Correlations of first semester marks with entrance measures

ENTRANCE MEASURES	A MARKS IN PROFESSIONAL COURSES									
	Schools									
	0	1	2	3	4	5	6	7	8	9
Terman	.38	.26	.15	.04	.40	.24	.35	.10	.20	.34
Tressler	.41	.20	.10	.07	.31	.23	.17	.21	.27	.32
Whipple	.32	.25	.14	.17	.36	.07	.19	.07	.04	.25
Teachers' Interest	.30	.28	.07	.20	.36	.25	.26	.12	.10	.28
Prognosis	.33	.34	.18	.05	.52	.18	.38	.21	.14	.34
High School Marks	.46	.43	.31	.09	.38	.17	.36	.13	—.04	.28
Habit Rating	.41	.31	.21	.18	.38	.08	.18	.05	.16	.21
Average	.37	.30	.17	.11	.39	.17	.27	.13	.12	.29
	B MARKS IN CONTENT COURSES									
Terman	.51	.37	.26	.31	.46	.38	.49	.25	.31	.25
Tressler	.42	.32	.35	.21	.38	.34	.53	.07	.40	.27
Whipple	.34	.22	.17	.25	.38	.24	.31	.14	.09	.18
Teachers' Interest	.42	.29	.19	.29	.25	.28	.33	.12	.20	.19
Prognosis	.52	.39	.30	.31	.49	.33	.45	.23	.37	.26
High School Marks	.57	.50	.43	.31	.53	.40	.55	.31	.18	.24
Habit Rating	.36	.23	.23	.13	.53	.12	.38	.03	.46	.27
Average	.45	.33	.28	.26	.43	.30	.43	.16	.29	.24

sixth semester correlations with content courses, however, there is not much difference in the degree of correlation, and School 3 yields slightly higher correlations than School 4.

TABLE 8

Correlations of second semester achievement test with entrance measures

ENTRANCE MEASURES	SCHOOLS									
	0	1	2	3	4	5	6	7	8	9
Terman	.66	.71	.54	.61	.75	.62	.71	.60	.68	.60
Tressler	.66	.51	.46	.58	.61	.55	.64	.48	.60	.46
Whipple	.51	.35	.36	.36	.51	.46	.38	.42	.37	.29
Teachers' Interest.	.54	.41	.33	.37	.49	.43	.41	.48	.42	.24
Prognosis	.72	.62	.52	.54	.71	.69	.69	.65	.60	.64
High School Marks	.39	.57	.53	.55	.63	.57	.54	.48	.34	.41
Habit Rating	.06	.22	.34	.19	.31	.33	.28	.28	.42	.23
Average	.51	.48	.44	.46	.57	.52	.52	.48	.49	.41

In comparing any school for first and sixth semester, however, we find apparently inexplicable inconsistencies. For example, in the first semester, professional marks are found to have a higher correlation with high school marks than with any other entrance measure in Schools 0, 1 and 2, and these correlations are, respectively, .46, .43 and .31. By the sixth semester, however, the correlations between professional marks and high school marks have dropped to .35, .12 and .09 and are now among the lowest correlations. On the other hand, in School 4, the correlation in the first semester was .38 and in the sixth semester had risen to .42, while in School 5 it rose from .17 to .43.

In general, first semester marks are correlated with entrance measures to a slightly higher degree than sixth semester marks, but this is not true for all schools (table 10). Again, the correlation tends to be slightly higher with content subjects than with professional subjects, although two schools are exceptions to this. Taking all schools together, the entrance measure which gave the highest correlation with normal school marks in the first semester was high school marks; in the sixth semester, the Terman Test.

The only entirely consistent result shown by any of these correlations is the lack of any relation between entrance measures and prac-

TABLE 9

Correlations of sixth semester marks with entrance measures

ENTRANCE MEASURES	A MARKS IN PROFESSIONAL COURSES									
	Schools									
	0	1	2	3	4	5	6	7	8	9
Terman	.38	.34	.12	.09	.39	.35	.23	.00	.51	.12
Tressler	.50	.30	—.04	.09	.38	.29	.34	.08	.37	.11
Whipple	.31	.22	.19	.15	.43	.24	.13	.07	.17	.14
Teachers' Interest	.43	.14	.19	.17	.48	.22	.21	.07	.35	.08
Prognosis	.43	.23	.23	.22	.48	.40	.23	—.04	.45	.08
High School Marks	.35	.12	.09	.12	.42	.43	.30	.03	—.01	.14
Habit Rating	.00	.09	.13	.25	.25	.26	.29	—.07	.13	.10
Average	.34	.19	.13	.16	.40	.31	.25	.02	.28	.11
	B MARKS IN CONTENT COURSES									
Terman	.49	.13	—.01	.24	.22	.29	.59	.20	.61	.11
Tressler	.39	.12	—.07	.32	.21	.26	.36	—.07	.22	.02
Whipple	.27	.39	.00	.25	.16	.20	.41	.03	.35	.07
Teachers' Interest	.51	.01	.09	.26	—.12	.11	.31	.02	.41	.09
Prognosis	.54	.19	.02	.18	.20	.32	.45	—.02	.45	.02
High School Marks	.44	.16	.06	.28	.35	.33	.37	.05	.23	.01
Habit Rating	.10	.18	.03	.24	.23	.26	.29	.16	.11	—.02
Average	.39	.17	.02	.25	.18	.25	.40	.05	.34	.04
	C PRACTICE TEACHING MARKS									
Terman	.33	.03	—.03	—.01	—.01	—.01	—.02	—.13	.29	.12
Tressler	.32	—.01	—.09	.05	.05	.05	—.03	.03	.37	.08
Whipple	.23	—.10	—.08	.26	—.05	—.03	.03	—.18	.12	—.04
Teachers' Interest	.22	.03	.02	.19	—.19	—.08	.15	—.08	.27	—.08
Prognosis	.29	.06	—.07	.04	—.11	.03	.01	—.05	.21	—.08
High School Marks	.45	.04	—.14	.05	.10	.02	.03	.08	.02	.29
Habit Rating	.02	.04	.04	—.16	.15	—.29	.21	.06	—.07	.03
Average	.27	.01	—.05	.06	.00	—.04	.05	—.04	.17	—.05

TABLE 10

Summary of correlations of entrance measures and normal school marks

CORRELATION OF ENTRANCE MEASURES	AVERAGE CORRELATION OF ALL SCHOOLS AND ALL ENTRANCE MEASURES	AVERAGE CORRELATION OF ALL ENTRANCE MEASURES FOR EACH SCHOOL									
		0	1	2	3	4	5	6	7	8	9
With first semester marks	.25	.39	.26	.22	.18	.33	.21	.30	.16	.19	.23
With sixth semester marks (excluding practice teaching)	.20	.37	.18	.07	.20	.29	.28	.32	.04	.31	.08
With practice teaching marks	.05	.27	.01	—.05	.06	.01	.04	.05	—.04	.17	—.05
With content subjects (1st and 6th)	.26	.42	.25	.15	.26	.30	.28	.42	.11	.31	.14
With professional subjects (1st and 6th)	.22	.36	.24	.15	.13	.40	.24	.26	.07	.20	.20
With second semester Achievement Test	.49	.51	.48	.44	.46	.57	.52	.52	.48	.49	.41

CORRELATION OF ENTRANCE MEASURES	AVERAGE CORRELATION OF ALL SCHOOLS FOR EACH ENTRANCE MEASURE						
	Terman	Tressler	Whipple	Teachers' Interest	Prognosis	High School Marks	Habit Rating
With first semester marks	.30	.28	.21	.24	.32	.33	.25
With sixth semester marks (excluding practice teaching)	.27	.19	.21	.20	.24	.21	.12
With practice teaching marks	.06	.08	.02	.05	.04	.09	.02
With content subjects (1st and 6th)	.32	.25	.22	.21	.30	.32	.22
With professional subjects (1st and 6th)	.25	.12	.20	.23	.26	.22	.15
With second semester Achievement Test	.65	.56	.40	.41	.64	.50	.27

tice teaching marks. The correlation is close to zero in every school except School 0, where it is .27, and it is close to zero with each of the entrance measures separately.

When an objective test of attainment is used instead of normal school marks, however, as was done in the second semester (table 8), the correlations are much higher and much more consistent from school to school. Judged by an objective test criterion, the Terman Test and the Prognosis Test yield fairly high correlations with the ability to do normal school work. But the low and inconsistent correlations with normal school marks and the lack of correlation with practice teaching marks leave the issue unsettled as to whether there is no relation between the entrance measures and teaching ability, or whether the normal school marks, both in practice teaching and in other subjects, are too ambiguous to afford a very satisfactory criterion.

Analysis of normal school marks. A study of the normal school marks may throw some light on the validity of the criterion. We have selected three aspects of the question which are related to the problem of student selection. These are: (1) the question of which students are eliminated, (2) the distribution of marks given in the different normal schools to all students and to those in the lowest tenth on entrance measures, and (3) a comparison of the marks given in first semester and in sixth semester to all students and to those in the lowest tenth on entrance measures. In the last case, the results of the objective achievement test given at the end of the second semester are also included.

Elimination could of course be determined with greater exactness from registrars' records than from the data we have. It is possible that there are unrecognized errors in the assumptions we are making here. We are assuming: (1) that practically all entering students took the entrance tests and that all those who took any of the tests did enter; (2) that the semester marks reported by the normal schools at the end of the first and sixth semesters included all of the tested groups who were still in normal school at the end of that semester.[1] If these assumptions are approximately correct, we can say that over 25 per cent of the students who entered left normal school before the end of the first semester and that approximately half of them left before the end of the sixth semester (See table 11). The normal schools vary greatly in their "holding power," if we might call it that. School 0, for example, gave practice teaching marks

[1] This assumption is known to be incorrect for two schools. (See table 11, footnote *a*)

TABLE 11

Survival of entering students at end of first and end of sixth semesters

NORMAL SCHOOL MARKS	PER CENT OF ENTERING STUDENTS TESTED WHO RECEIVED SEMESTER MARKS										
	Schools										
	0	1	2	3	4	5	6	7[a]	8[a]	9	Total
First semester											
Content	100	87	68	76	75	96	92	53	31	92	73
Professional	99	87	37	73	72	76	92	52	31	83	70
Sixth semester											
Content	79	41	50	56	27	70	44	35	21	51	46
Professional	70	41	58	64	50	73	54	34	21	46	51
Practice teaching[b]	82	55	61	66	27	77	60	40	21	54	54
No. of students with entrance tests	71	261	249	191	210	202	178	161	141	213	1 867

a These two schools did not report semester marks for students in special courses and these figures do not represent the true survival, which is a little over 60 per cent in the sixth semester. First and sixth semesters are comparable, however.

b Marks given in fifth or sixth semester.

(in fifth or sixth semester) to 82 per cent of the entering group, School 4 to only 27 per cent. This study covered the years of transitional requirements for elementary teachers and many students undoubtedly left to teach on a "limited normal certificate" after two years, while others with incomplete training returned to complete their normal school work. The number of graduates in 1930 was larger than the number of 1927 entering students in all but two schools.[1]

When we compare the amount of survival of all students in the sixth semester with the survival of those in the lowest tenth on each of the entrance measures (compare table 12 and table 11), we find only two schools (Schools 0 and 9) in which the average percentage of the lowest tenth in various entrance measures which completed sixth semester is to any extent smaller than the percentage of the whole number of students. If we compare the percentage surviving in sixth semester in each school with the percentage of the lowest tenth on the separate entrance measures surviving in the same school, School 9 seems to be the only one in which there appears to be any consistent tendency for elimination to occur more often among the students of the lowest tenth. We find very great inconsistencies between schools in regard to separate tests. In School 4, for example, 87 per cent of the lowest tenth on Terman survive, while in School 9, the percentage is only 25. In School 7, 100 per cent of the lowest tenth on Habit Rating survive, and in School 9, 23 per cent. Elimination does not occur more often among the lowest on one test than on another; in fact, it seems entirely unrelated to the tested abilities of the students.

When we turn to the distribution of the marks given in normal school, we find an explanation for the lack of correlation with the entrance tests. Less than a quarter of the marks given in any school are other than B or C (table 13). With so small a range of marks, correlations will necessarily be low. We find that schools also vary in their practice of giving very low or very high marks. School 4 gives the smallest per cent of A's and the largest per cent of E's. In Schools 2, 3 and 8 we find almost no failing marks, while Schools 0, 1 and 4 give 2 per cent or more of failing marks.

Table 14 may be compared with table 13, to see whether the students who were in the lowest tenths on the entrance measures tended to receive high marks less frequently and low marks more

[1] W. W. Coxe and E. L. Cornell. Pre-Service and In-Service Training of Elementary School Teachers. Educational Monograph of New York State Teachers Association, No. 2, 1932. (See table 10, p. 17)

TABLE 12

Survival among those in lowest tenth of each entrance measure

ENTRANCE MEASURES	PER CENT OF LOWEST TENTH ON EACH ENTRANCE MEASURE WHO RECEIVED SIXTH SEMESTER MARKS										
	Schools										
	0	1	2	3	4	5	6	7	8	9	Total
Terman	69	35	46	70	87	35	50	50	40	25	61
Tressler	88	53	54	54	62	65	43	33	25	39	52
Whipple	93	47	64	64	62	71	47	60	50	36	58
Teachers' Interest	80	50	67	63	69	68	35	67	57	52	61
Prognosis	69	47	67	56	56	76	69	46	75	17	59
High School Marks	.....	50	56	78	63	75	53	70	67	24	54
Habit Rating	75	64	44	57	39	77	45	100	20	23	49
Average	68	50	57	63	62	67	49	61	46	31	56

TABLE 13

Percentage distribution of all marks given in each normal school

MARKS	SCHOOLS									
	0	1	2	3	4	5	6	7	8	9
A	6.6	8.3	8.4	11.4	5.0	5.3	14.9	10.0	5.7	6.4
B	33.8	36.3	44.2	38.0	32.2	36.4	37.4	42.5	44.8	40.8
C	43.6	41.0	42.4	44.2	47.7	50.6	32.7	39.2	39.2	37.7
D	14.1	12.4	4.9	6.1	12.8	6.3	14.1	6.9	9.6	13.4
E	2.0	2.0	0.1	0.3	2.3	1.4	0.8	1.4	0.6	1.6
Total number of marks	305	812	680	642	525	787	609	346	176	692

TABLE 14

Average of percentages of those in the lowest tenth on each entrance measure who received highest and lowest marks in each school

	SCHOOLS									
	0	1	2	3	4	5	6	7	8	9
Marks of A	.3	2.7	3.6	5.1	2.8	1.7	3.8	4.0	0.5	5.3
Marks of E	4.2	5.9	0.2	....	5.5	2.4	4.4	2.3	0.5	5.5

frequently than all students. This is generally true, although the difference between the poorest students on tests and all students is not very marked and there are occasional exceptions. In School 3, for example, there were no marks of E among the lowest tenths on tests, although some marks of E were given in the school. In general, the marks of the lowest tenths are lower than the average, but we nevertheless find some A's among the lowest tenth, even though only a very small percentage of all students received marks of A.

There is some evidence that the marks given in the first semester and in the sixth semester are not comparable. At least it is clear that more low marks than high marks were given in the first semester, while the opposite is the case in the sixth semester (table 15). This tendency is true for almost every school in both content and professional subjects, in spite of great differences in the actual percentages in the various schools. It is true for practice teaching in all but one school. School 5 is exceptional in this respect: It gave A's to only 1 per cent of students in practice teaching and marks of D or E to 7 per cent. Three schools gave no practice teaching marks below C.

It must be remembered, however, that while these students include all students in the first semester, they represent only about half of the total number of sixth semester students. They may be on the whole the better students, but it would be very unlikely that the distribution of the marks of the other half of sixth semester students could so weight the poorer end as to make the percentages of high and low marks comparable in both semesters.

Table 16 gives a comparison of the low and high marks received by students who were in the lowest tenth on certain of the entrance measures. The percentages of all students, from table 15, are included for comparison. In the first semester, there are fewer A's and more D's or E's among the lowest tenth on each of the five measures reported than among all students. This is true also in the sixth semester except on Habit Rating, although there is much less difference between the lowest tenth and all students in the sixth semester than in the first. It is interesting to note that although none of the lowest tenth on high school marks received A's in content subjects in the first semester, 11.4 per cent of them received A's in practice teaching. The per cent of the lowest tenth on each entrance measure which fell below the thirtieth percentile and exceeded the seventieth percentile on the second semester achievement test is also included in the table. If we can say that the lowest 30 per cent on

Table 15

High and low marks given to all students in first and sixth semester

MARKS IN COURSES IN EACH SEMESTER	PER CENT OF ALL STUDENTS RECEIVING MARKS OF SPECIFIED TYPE										TOTAL
	Schools										
	0	1	2	3	4	5	6	7	8	9	
A First semester											
Content	6	3	6	3	4	6	12	1	2	8	5.5
Professional	6	8	2	2	3	4	14	8	7	1	5.6
D, E First semester											
Content	20	16	5	9	15	5	24	16	14	23	14.5
Professional	36	32	12	9	32	20	18	15	21	29	23.2
A Sixth semester											
Content	11	8	15	18	7	9	20	9	3	8	11.5
Professional	8	8	8	14	8	7	16	25	10	3	9.9
D, E Sixth semester											
Content	9	1	2	7	2	2	19	..	10	5	4.9
Professional	10	3	3	2	5	3	6	2	..	3	3.7
A Practice teaching	3	17	10	24	5	1	17	11	7	12	11.6
D, E Practice teaching	..	3	5	6	2	7	..	2	..	1	3.2
No. of students involved											
First semester											
Content	71	226	170	145	157	193	164	85	43	195	1 449
Professional	70	227	91	140	152	154	164	84	43	177	1 302
Sixth semester											
Content	56	108	124	107	56	141	79	57	30	108	866
Professional	50	107	144	123	104	147	96	55	30	97	953
Practice teaching	58	144	151	127	56	156	106	65	30	115	1 003

TABLE 16

High and low marks received by lowest tenth of students on certain entrance measures in first and sixth semesters and on second semester achievement test

MARKS IN COURSES IN EACH SEMESTER	PER CENT OF THOSE IN LOWEST TENTH ON					PER CENT OF ALL STUDENTS
	Terman	Teachers' Interest	Prognosis	High School Marks	Habit Rating	
A First semester						
Content	0.6	0.6	0.7	....	2.8	5.5
Professional	0.7	1.3	2.9	1.7	....	5.6
D, E First semester						
Content	26.4	29.8	27.5	30.3	27.3	14.5
Professional	37.0	34.0	34.8	39.4	37.4	23.2
A Sixth semester						
Content	6.8	5.6	7.9	7.7	2.9	11.5
Professional	1.2	1.7	....	1.4	5.3	9.9
D, E Sixth semester						
Content	6.8	6.5	10.1	10.7	4.4	4.9
Professional	8.4	6.1	6.4	5.4	7.9	3.7
A Practice teaching	8.2	6.8	8.7	11.4	5.1	11.6
D, E Practice teaching	3.5	4.3	1.1	2.9	2.5	3.2
Second semester achievement test						
Per cent in highest three-tenths	....	11.0	2.3	5.9	15.7	30
Per cent in lowest three-tenths	76.5	57.5	71.5	68.4	53.0	30
Number of students in lowest tenth	234	195	170	138	170	....

the achievement test are doing below average work, then it is apparent that about twice as many of the students lowest on entrance tests were doing poor work in the second semester as had received marks below C in the first semester. If the entrance tests measure ability, therefore, and if the second semester achievement test measures achievement, then semester marks do not appear to be an adequate measure of either. If the tests are not a valid criterion by which to judge the marks, nevertheless the variation between schools and between semesters in the marks given makes it evident that the marks themselves are not comparable measures and therefore equally invalid at least for comparing students from different schools.

We can therefore say that while the abilities of students as judged by entrance tests seem to have little relation to normal school marks, this may be evidence of lack of validity of the marks rather than of the tests. We turn therefore to our final section, a study of the entrance tests as related to teaching success.

HOW THE ENTRANCE TESTS WERE RELATED TO THE RATING OF TEACHING SUCCESS

It was stated at the outset that our results would be disappointing since we were unable to secure a highly valid criterion of teaching success. We have referred in the introduction to the various kinds of ratings we have on these teachers. The Ten-Point Rating (on a scale we devised for this purpose) was secured from the supervisors of approximately 550 teachers who had been tested at entrance to normal school. The other ratings were limited to a selected group of 112 teachers who were visited by two expert judges and rated on two published rating scales for teachers. These teachers were also rated on one of these scales by their supervisors and were given a so-called "general rating" by them (that is, *above average, average* or *below average*). We have therefore one rating of a fairly large group, which can be checked by comparison with other ratings of a rather small group. The small group was selected with considerable care in order to obtain a representative sampling from each normal school and actually does give us a fair sampling of the total range of tested ability, with some minor fluctuations. The percentage of the visited group above and below the medians of all tested students is shown by the following:

	Terman	*Tressler*	*Prognosis*	*Second semester achievement*
Above median	52.8	44.7	58.9	52.5
Below median.	47.2	55.3	41.1	47.5

Prognostic value of entrance tests as indicated by correlations with the Ten-Point Rating. We may consider first the rating of teaching of the larger group on the Ten-Point Rating Scale. We have two ratings of teachers on this scale, one made at the end of their first year of teaching, the other at the end of their second year. The correlations of this scale with the entrance measures are shown in table 17. They are all approximately zero with the exception of Habit Rating, which reaches the order of $r = .1$. But we observe that achievement in normal school and even practice teaching marks are not much better indications of teaching success as judged by the Ten-Point Rating. Either tested ability and normal school marks do not prognosticate teaching success or the Ten-Point Rating is not a valid criterion of teaching success.

Comparison of various ratings of teachers. We can, fortunately, check the ten-point rating by several other ratings on the smaller group which was visited by two expert raters who observed the teacher for a period and rated her on two scales. One scale was the *Torgerson Diagnostic Teacher Rating Scale of Instructional Activities*, which attempts to evaluate characteristics of the teaching technic. The other was the *Almy–Sorenson Rating Scale for Teachers*, which is a rating of personal qualities or personality traits regarded as important in teachers.[1] Each of the two raters made his or her rating independently of the other, but compared ratings and sometimes changed certain items in the light of combined observations. The raters did not compromise upon items on which they disagreed, however, if discussion did not change their opinion. This procedure ought to give about as similar ratings of an individual as could ever be obtained from two judges, but the correlations between the ratings of the two judges on the same scale are not so high as we might expect to get by this method. They are:

Torgerson, 2 raters	$r = .910$
Almy-Sorenson, 2 raters	$r = .506$

The difference in the size of these correlations indicates clearly that much better agreement can be expected as to the technic of conducting a recitation, when the judges observe the same period, than as to the teacher's personality under the same conditions.

This result was anticipated, since it is difficult to judge personal qualities without longer acquaintance. We therefore had the supervisors of these teachers (the same supervisors who used the Ten-

[1] For further discussion of these ratings see Pre-Service and In-Service Training of Elementary School Teachers, Chap. V.

TABLE 17

Correlations of ten-point rating of 550 teachers by supervisors with entrance measures and normal school work

	TEN-POINT RATING OF TEACHING	
	First Year	Second Year
Entrance measures		
Terman	—.06	+.01
Tressler	.00	+.06
Whipple	—.02	+.04
Teachers' Interest	—.01	+.07
Prognosis	—.03	+.03
Habit Rating	+.11	+.13
Normal school work		
Second semester achievement	—.03	—.01
Practice teaching	+.13	+.21
Number of cases	467 to 544	364 to 417

Point Rating Scale) also rate them on the Almy-Sorenson Scale. The supervisors agreed in their own ratings on the two different scales much better than they agreed with the visiting raters on the same scale, as is shown by the correlations.

Almy-Sorenson Scale, supervisors and raters..... $r = .37$
Almy-Sorenson Scale and Ten-Point Rating, supervisors.. $r = .67$

These correlations are susceptible of more than one interpretation. They might mean that the visiting raters did not know the teachers well enough to give reliable ratings, which is probably true to some extent. They might mean, on the other hand, that when supervisors make a judgment of general teaching success (as on the Ten-Point Scale) they judge more in terms of personality traits than of teaching technics. We might have determined whether this were true if the supervisors had also used the Torgerson Scale, but unfortunately, they did not. We can, however, compare the Almy-Sorenson and the Torgerson ratings made by the visiting raters with the Ten-Point Scale. If the Ten-Point ratings should be based largely on personality judgments, we should expect that the correlation of this scale with the Almy-Sorenson Scale would be higher than its correlation with the Torgerson Scale.

The correlations are as follows:

Almy-Sorenson (raters) and Ten-Point Rating (supervisors) ..... r = .48

Torgerson (raters) and Ten-Point Rating (supervisors) ..... r = .41

The difference is in the direction indicated but it is not very large. This, however, might be partly due to the lower reliability of the Almy-Sorenson ratings as compared with the Torgerson.

When the raters' and supervisors' judgments on Almy-Sorenson were combined, the correlation was considerably higher, both with the Ten-Point Scale and with the Torgerson Scale. In this case, the correlation was higher with the Torgerson Scale.

Almy-Sorenson (combined) and Ten-Point Rating... r = .65

Almy-Sorenson (combined) and Torgerson..... r = .78

The Ten-Point Rating was, as has been stated before, used by the supervisors in two successive years. The agreement between the ratings of the same individuals in the first and second year was, for 422 cases:

Ten-Point Rating, 1931 and 1932..... r = .55

Considering that we should expect teachers to change their relative positions somewhat between the end of their first and second year of experience, this correlation seems to indicate a fair amount of consistency in supervisors' opinions.

TABLE 18

Correlations of other measures of teaching success of 112 teachers with entrance measures and normal school work

	ALMY-SORENSON SCALE	TORGERSON SCALE	COMPOSITE CRITERION
Entrance measures			
Terman	+.09	+.10	+.03
Tressler	+.09	+.10	+.12
Whipple	+.10	+.08	+.18
Teachers' Interest	+.09	+.09	+.10
Prognosis	+.13	+.13	+.08
High School Marks.....	—.05	.00	+.08
Habit Rating	—.04	—.05	—.05
Normal school work			
Second semester achievement.	+.10	+.01	+.10
Practice teaching	.....	.....	+.28

The Almy-Sorenson and Torgerson ratings of the first year were also correlated with the Ten-Point Ratings of the second year. The correlations were naturally lower for the second year than for the first, but that of the second year's Ten-Point Rating with the Almy-Sorenson was nevertheless higher than the first year's Ten-Point Rating with the Torgerson.

Almy-Sorenson (combined) and Ten-Point Rating, 1931	$r = .64$
1932	$r = .46$
Torgerson and Ten-Point Rating, 1931	$r = .41$
1932	$r = .32$

This discussion of the intercorrelations of the various rating devices leaves us with the impression that we have no criterion of teaching success in which we can place much confidence. We may add, however, that the ratings do not seem to have any less validity than the normal school marks.

Prognostic value of entrance tests as indicated by correlations with Almy–Sorenson and Torgerson ratings. Although our criterions were disappointing, we investigated the relation of the entrance measures to the ratings made by the visiting raters. The correlations of the careful ratings made by a small selected group of experienced raters on carefully standardized rating scales are no more significant than those made by a large number of supervisors on the admittedly rough Ten-Point Scale. (See tables 18 and 17.) We therefore thought it worth while, since one measure of teaching success seemed no better than another, to develop a composite criterion of all the measures we had, to learn whether this would yield higher correlations or could be used to obtain a multiple correlation high enough to have any prognostic value. The results were negative, as the correlations of the entrance tests with the composite criterion[1] show (table 18). The multiple correlation of the seven entrance measures with the composite criterion yielded the value $R = .263$, which corrected for attenuation, became $R = .31$.

Prognostic value of combined low scores on the entrance tests. It was pointed out in the first section of this report that we could

[1] The composite criterion was made up of seven items: (1) decile standing on Almy-Sorenson (combined); (2) decile standing on Torgerson; (3) Ten-Point Rating; (4) index of difficulties of discipline; (5) index of difficulties of organization; (6) index of difficulties of routine; (7) index of difficulties of understanding children. These indices were derived from the difficulties reported both by the teachers themselves and by their supervisors. See *Pre-Service and In-Service Training.*

probably select the poorest *students* among the entrants by taking those who were in the lowest 10 per cent on at least three of the entrance measures. There were nine of these students in the *visited-rated* group of 112 teachers. Table 19 shows how they stood in normal school and on the various rating scales. Seven out of eight measured were below the thirtieth percentile on the second semester achievement test and none were above the seventieth percentile in any of the normal school measures. It would therefore be correct to say that these were among the poorest students both in ability and in accomplishment. In actual teaching situations, however, some were judged to be among the best, and about half were above

TABLE 19

Percentile standing in normal school marks and teacher ratings of nine students who were in lowest tenth on at least three of the seven entrance measures

PERCENTILE RANK	NORMAL SCHOOL MEASURES			
	Second Semester Achievement	Sixth Semester Content	Sixth Semester Professional	Practice Teaching Marks
Above 50th	..	2	..	A
Below 50th	8	5	8	B 3
				C 4
Above 70th	..	..	..	D
Below 30th	7	3	6	E
Number	8	7	8	Number 7

PERCENTILE RANK	RATING OF TEACHING SUCCESS				
	Almy-Sorenson	Torgerson	Ten-Point 1931	Ten-Point 1932	Composite Rating
Above 50th	4	3	2	1	3
Below 50th	5	6	3	3	3
Above 70th	1	3	2	1	..
Below 30th	3	3	1	3	3
Number	9	9	5	4	6

the median on the various scales. Apparently there are elements of successful teaching which have not been touched upon by the measures we have used.

WHAT THE CRITERIONS MEASURE

Validity of the criterions as judged by differences between good and poor teachers selected by these criterions. Assuming the validity of any of the criterions, the data indicate that high school marks, intelligence tests, achievement tests and normal school marks do not have much relation to teaching success. Possibly, however, the difficulty is with the criterions. Perhaps we have not succeeded in measuring success with sufficient refinement to yield high correlations with other measures. Perhaps if we should select the best and poorest teachers on the basis of our criterions, we should find that there were measurable differences between them. Whether we find prognostic indications or not, however, we may be able to find indications, through an analysis of the differences, as to what the criterions do measure and whether they have any validity.

We have several criterions by which we may select the best and poorest teachers. Perhaps the criterion least open to objection is the opinion of the local supervisor (whether superintendent, principal or grade supervisor) as to whether the teacher is above average, average or below average. At least it is this sort of opinion which is the determining factor in regard to employment and promotion and is commonly accepted as valid. We have such a "general rating" for 105 of the group which was visited and rated. Of these,

TABLE 20

Percentage of "above average," "average" and "below average" teachers having high, medium and low average standings on entrance tests

AVERAGE PERCENTILE RANK ON FIVE ENTRANCE TESTS	GENERAL RATING		
	Above Average	Average	Below Average
In highest 30 per cent	16.0	13.1	22.2
In middle 40 per cent	58.0	56.5	22.2
In lowest 30 per cent	26.0	30.4	55.5
Number of cases	50	46	9

TABLE 21

Percentage of "above average," "average" and "below average" teachers having high, medium and low average standings in sixth semester marks and practice teaching

NORMAL SCHOOL MARKS	SIXTH SEMESTER CONTENT COURSES			PRACTICE TEACHING		
	General Rating			General Rating		
	Above Average	Average	Below Average	Above Average	Average	Below Average
A.....	12.0	2.2	....	10.0	15.2	....
B or C.....	60.0	65.3	66.6	78.0	69.5	100.0
D.....	....	6.5	....	....	4.4	....
No mark	28.0	26.1	33.3	12.0	10.9	....
Number of cases	50	46	9	50	46	9

TABLE 22

Percentage of "above average," "average" and "below average" teachers having high, medium and low composite ratings

PERCENTILE RANK ON COMPOSITE RATING	GENERAL RATING		
	Above Average	Average	Below Average
In highest 30 per cent.....	45.3	11.9	....
In middle 40 per cent.....	52.4	45.3	....
In lowest 30 per cent.....	2.4	42.9	100.0
Number of cases.....	42	42	8

50 were rated "above average," 46 "average," and 9 "below average." It is evident that "below average" was interpreted as definitely unsatisfactory. We have used this general rating, first for a brief comparison with entrance tests and normal school marks, and then as a check on the composite criterion. Tables 20 and 21 show how this rating is related to the entrance tests and to normal school work, and table 22 shows its relation to the composite criterion. Neither work in normal school nor the entrance tests show much relation to this general rating. The only marks below C in either

content courses or practice teaching were given to persons later rated as "average" teachers. A fair percentage of A's in practice teaching were given to those who were rated as "average." None of the very small number of "below average" teachers received normal school marks below C. Some of the "above average" teachers had an average rank on entrance tests in the lowest 30 per cent and some of the "below average" teachers were in the highest 30 per cent.

Table 22 shows that all those rated as "below average" were in the lowest 30 per cent on the composite rating; those rated "average" were not very likely to be among the highest 30 per cent; and those rated "above average" were not very likely to be among the lowest 30 per cent. It would seem that those rated "above average" and also in the highest 30 per cent on the composite rating might reasonably be regarded as the best teachers and that those "below average" and in the lowest 30 per cent on the composite rating might be regarded as the poorest.

Our first selection of teachers will then concern the 19 individuals whom we may call the "best" and the eight individuals whom we may call the "poorest" teachers. What differences are there between these two groups?

Differences between "best" and "poorest" teachers. Tables 23 to 28 show the differences we have chosen for analysis for these two groups. From scrutiny of table 23 we may say that the "best" and "poorest" teachers are not distinguished adequately, if at all, by any of the entrance measures. It is perhaps worth while to point out that a large majority of the "poorest" teachers were above the median in high school marks, more than half of them were above the seventieth percentile and none were below the thirtieth percentile. The only test in which we find none of the "poorest" teachers above the seventieth percentile is the Whipple Reading Test. There is no test in which some of those in the lowest 30 per cent were not later found among the best teachers. Thus, even when we have a criterion which is as valid as we can make it, we do not find that the entrance tests discriminate between even the "best" and the "poorest" teachers.

From table 24 we may discover the relations between the "best" and "poorest" teachers and the work in normal school. There is not much difference between the two groups in the ratings in the objective achievement test, in the marks in content courses and professional courses, nor in practice teaching. Possibly there is a tendency for those who will become the "poorest" teachers to receive low marks

TABLE 23

Percentage of "best"[1] and of "poorest"[1] teachers having good and poor standing on entrance measures

PERCENTILE STANDING	TERMAN		TRESSLER		WHIPPLE		TEACHERS' INTEREST		PROGNOSIS		HIGH SCHOOL MARKS		HABIT RATING	
	Best	Poorest	Best	Poorest	Best	Poorest	Best	Poorest	Best	Poorest	Best	Poorest	Best	Poorest
Above 50th	68.4	50.0	47.4	50.0	52.6	37.5	73.6	75.0	63.2	37.5	63.2	85.7	33.3	57.1
Below 50th	31.6	50.0	52.6	50.0	47.4	62.5	26.3	25.0	36.8	62.5	36.8	14.3	66.7	42.9
Above 70th	52.6	12.5	31.6	37.5	26.3	....	63.2	62.5	42.1	37.5	36.8	57.1	16.7	14.3
Below 30th	15.8	25.0	26.3	50.0	21.0	50.0	15.8	25.0	21.0	50.0	10.5	....	27.8	14.3
Number of cases.	19	8	19	8	19	8	19	8	19	8	19	7	18	7

[1] Explanation of method of selection in text.

TABLE 24

Percentage of "best"[1] and of "poorest"[1] teachers having good and poor standing on normal school measures

PERCENTILE STANDING	SECOND SEMESTER ACHIEVEMENT TEST		MARKS IN SIXTH SEMESTER				PRACTICE TEACHING		
			Content		Professional				
	Best	Poorest	Best	Poorest	Best	Poorest	Marks	Best	Poorest
Above 50th	73.4	50.0	53.8	80.0	52.6	28.6	A or B	64.7	75.0
Below 50th	26.6	50.0	46.2	20.0	47.4	71.4	C, D or E	35.3	25.0
Above 70th	40.0	37.5	23.1	....	31.6	14.3	A	23.6	....
Below 30th	....	12.5	38.5	20.0	47.4	71.4	D or E	....	....
Number of cases	15	8	13	5	19	7		17	8

[1] Explanation of method of selection in text.

TABLE 25

Percentage of "best"[1] and of "poorest"[1] teachers receiving high and low ratings on different rating scales

PERCENTILE STANDING	ALMY-SORENSON		TORGERSON		TEN-POINT RATING	
	Best	Poorest	Best	Poorest	Best	Poorest
Above 50th	89.5	....	89.5	....	84.2	....
Below 50th	10.5	100.0	10.5	100.0	15.8	100.0
Above 70th	63.5	....	42.1	....	84.2	....
Below 30th	....	100.0	10.5	50.0	....	100.0
Number of cases	19	8	19	8	19	5

[1] Explanation of method of selection in text.

TABLE 26

Percentage of "best"[1] and of "poorest"[1] teachers possessing necessary personal qualifications, according to supervisors

	BEST	POOREST
Has necessary personal qualifications.....	94.7	12.5
Has not necessary personal qualifications.....	....	50.0
No reply or doubtful.....	5.3	37.5
Number of cases	19	8

[1] Explanation of method of selection in text.

TABLE 27

Percentage distribution of age and sex of "best"[1] and "poorest"[1] teachers

SEX	BEST	POOREST	AGE AT NORMAL SCHOOL ENTRANCE	BEST	POOREST
			16	25.0	12.5
Female	89.5	100.0	17	25.0	....
Male	10.5*a*	....	18	25.0	50.0
			19	25.0	37.5
Number of cases...	19	8	Number of cases	16	8

[1] Explanation of method of selection in text.

a This is 28.6 per cent of the total number of males rated. Seven of the 112 teachers rated were males. One was rated *below average*, but had no composite rating.

Table 28

Percenatge of "best"[1] and of "poorest"[1] teachers having certain types of difficulty in teaching, according to supervisors

Amount of difficulty reported by supervisor	Type of difficulty									
	Discipline		Organization		Routine		Personal relations		Understanding children	
	Best	Poorest	Best	Poorest	Best	Poorest	Best	Poorest	Best	Poorest
Large proportion of type of difficulty	12.5	100.0	....	33.3	....	16.7	....	66.7	....	50.0
Small proportion of type of difficulty	12.5	....	37.5	50.0	18.7	33.3	....	33.3	25.0	50.0
None	75.0	....	62.5	16.7	81.3	50.0	100.0	....	75.0	....
Number of cases	16	6	16	6	16	6	16	6	16	6

[1] Explanation of method of selection in text.

in professional courses, but the " best " teachers also receive more low marks than high in professional courses.

Table 25 shows the percentage of the " best " and of the " poorest " teachers rated high and low on the separate rating scales. The Almy-Sorenson and the Ten-Point Rating approximate more closely than the Torgerson the criterion of best and poorest. This suggests that the technic of teaching is less important than personality factors in determining what is commonly meant by very good or poor teachers. Furthermore, the correspondence, with all three scales, is closer for the " poorest " teachers than for the " best," which suggests that there may be more agreement as to what constitutes very poor teaching than very good. Possibly, however, it only means more caution in expressing one's opinion when it is adverse.

A little further light is thrown on the question of personality from other data obtained from the supervisors for Pre-Service and In-Service Training of Teachers which involved these same teachers. The supervisor was asked to indicate, on a questionnaire blank dealing primarily with teachers' difficulties, whether or not the teacher possessed the personal qualifications that would enable her to become a good teacher. Table 26 indicates that the best teachers were almost invariably judged to possess them, while only one out of the eight poorest teachers was so judged.

The questionnaires referred to in the last paragraph furnished other data, many of which have been analyzed for a larger group which will be discussed presently, and some of which we have also presented for the group under immediate consideration. Table 27 shows the age and sex distribution of the " best " and " poorest " teachers. There are too few males in the group to make the sex distribution of much consequence, but it is interesting to note that in the age distribution, which ranged from 16 to 19, a much larger per cent of the " poorest " teachers than of the " best " were 18 or 19 years old when they entered normal school. This would suggest that intelligence has some influence in determining good teaching, since younger high school graduates tend to be more intelligent than older ones, yet we have been unable to find any very clear-cut evidence that intelligence is of much significance, at least at the levels at which it is found in teacher-training institutions.

In table 28, however, we come upon some evidence that may afford constructive suggestions. The questionnaire sent to the supervisors had a list of 29 difficulties most often met by beginning teachers, from which the supervisors were asked to check those that applied.

The difficulties were later grouped into those relating to discipline, organization of work, routine matters such as record-keeping, personal relations, and the understanding of children. We then determined the number of difficulties of each type checked for each teacher and its relation to the total number of that type. Table 28 shows the percentage of the "best" and of the "poorest" teachers who were reported to have a large proportion of each type of difficulty (60 per cent or more of the type), a small proportion, or none. We find that more of the "best" teachers than of the "poorest" are reported to have no difficulties of each type, and similarly more of the "poorest" teachers have a large proportion of difficulties of each type. Especially do difficulties of discipline, of personal relations, and of understanding children differentiate between them. Here then is some evidence of what our criterions are measuring.

The number of cases dealt with in this group of "best" and "poorest" teachers was so small that we felt we should like to check these findings for a larger group. We therefore turned to the Ten-Point Rating Scale in order to have a wider selection of teachers from which to select good and poor.

Differences between "good" and "poor" teachers. We selected from the percentile distribution of the Ten-Point Ratings, as nearly as we could, the highest and lowest 30 per cent. Unfortunately, we could not do this exactly, as the distribution was skewed; and the better group includes, as before, more teachers than the poorer group. The "good" teachers include approximately the highest 30 per cent, while the "poor" teachers comprise only 20 per cent at the low extreme.

Table 29 shows the per cent of "good" and of "poor" teachers whom the supervisors believed to possess or not to possess the necessary personal qualifications. The difference is by no means so

TABLE 29

Percentage of "good" and "poor" teachers possessing necessary personal qualifications

	GOOD	POOR
Has necessary personal qualifications	100.0	79.3
Has not necessary personal qualifications	....	10.9
No reply or doubtful	....	9.9
Number of cases	154	101

clear cut as for the smaller group of "best" and "poorest" teachers (see table 26), but neither is the distinction between "good" and "poor" so well established as that between "best" and "poorest."

Table 30 may be considered with table 28 to compare types of difficulty reported. The same tendencies which were found for the small group are shown by the larger group, although again with rather less sharp distinctions. Difficulties of personal relations, for the larger group, do not indicate so marked a contrast as for the smaller group, but difficulties of discipline and of understanding children show a decided contrast in the large as well as in the small group.

Furthermore, we may discover whether "good" and "poor" teachers are distinguished by the difference in the improvement they make during their first year of teaching. If so, this will furnish a further check on the validity of the classification. Table 31 shows how the "good" and "poor" teachers differ in respect to showing improvement and *marked* improvement. We find a larger percentage of "good" teachers than of "poor" for whom no reply is made, which indicates in most cases that no difficulties were mentioned by the supervisor and therefore improvement could not be measured. Owing to this fact, the "poor" teachers make some improvement oftener than do the "good" teachers, but when *marked* improvement is considered, the opposite is the case. Nearly half of the "poor" teachers make no marked improvement, while only 13 per cent of the "good" teachers are in this category. These findings are consistent with what we should expect of "good" and "poor" teachers and tend thus to indicate that the rating has some validity.

Another item on which we have information from the supervisors is one concerning general causes of weakness in these teachers. The comparison of "good" and "poor" teachers on this item is made in table 32. Quite clearly, supervisors tend to find most of the weaknesses of "good" teachers to be lack of experience, while those of "poor" teachers are scattered over a wider range of causes, including notably lack of understanding of children, lack of a good professional attitude, and traits that prevent good discipline.

Another item in which we have found significant differences between "good" and "poor" teachers is in the grade taught or teaching position (table 33). Fewer of the "poor" teachers are found in kindergarten and primary positions and many more in rural schools ("all grades"). The interpretation of this may not be quite so obvious as it seems. At first thought, it appears probable that kindergarten and primary teachers may be better selected as to

TABLE 30

Percentage of "good" and of "poor" teachers having certain types of difficulty in teaching, according to supervisors

AMOUNT OF DIFFICULTY REPORTED BY SUPERVISORS	TYPE OF DIFFICULTY									
	Discipline		Organization		Routine		Personal Relations		Understanding Children	
	Good	Poor	Good	Poor	Good	Poor	Good	Poor	Good	Poor
Large proportion of type of difficulty...	11.0	53.5	2.6	28.7	5.2	10.9	1.3	7.9	5.8	31.7
Small proportion of type of difficulty..	15.6	24.7	39.6	48.5	18.2	43.6	6.5	19.8	33.8	42.6
None	73.5	21.8	57.8	22.8	76.6	45.5	92.2	72.3	60.4	25.7
Number of cases	154	101	154	101	154	101	154	101	154	101

TABLE 31

Percentage of "good" and of "poor" teachers reported to have made improvement in teaching difficulties

	GOOD	POOR		GOOD	POOR
Improvement	67.5	80.2	Marked improvement	56.5	35.7
No improvement	2.0	4.0	No marked improvement	13.0	48.5
No reply	30.5	15.8	No reply	30.5	15.8
Number of cases	154	101	Number of cases	154	101

aptitude for the job during their training period, and that of the others, the poorest candidates at graduation are least likely to be selected for city or village jobs and drift into the less desirable one-

TABLE 32

Percentages of total number of weaknesses mentioned by supervisors for "good" and for "poor" teachers which are ascribed to certain causes

CAUSE OF WEAKNESS	GOOD	POOR
Lack of experience	67.3	32.2
Inadequate general education	0.9	4.9
Lack of understanding of children	2.8	12.0
Lack of personal contacts outside	5.6	9.4
Lack of good professional attitude	....	5.6
Assignment to work not prepared for	8.4	8.2
Inadequate teacher training	2.8	5.3
Lack of forcefulness, initiative	....	2.2
Inadequacies of personality	....	2.6
Difficulty with discipline	....	2.2
Miscellaneous	12.1	15.4
Total number of weaknesses mentioned	107	267
Per cent of teachers for whom some cause given	55.2	92.1

TABLE 33

Percentage of "good" and of "poor" teachers in various teaching positions

TEACHING POSITION	GOOD	POOR
Kindergarten	11.8	2.6
Primary (1–3)	38.2	20.5
Intermediate (4–6)	22.6	24.4
Advanced (7–8)	5.9	12.8
All grades	6.9	24.4
Subject teachers	11.8	11.5
Miscellaneous	2.9	3.8
Number of cases	102	78

room schools. While this is probably part of the explanation, there may be other factors, such as the lack of help from supervisors in one-room schools, the lack of necessary equipment, the greater diffi-

culty of teaching a wide range of grades, especially without experience, the lack of stimulation and opportunity for observation of other teachers in the rural community and so forth, all of which might tend to induce a " poor " rating of a teacher in a rural school, even though the same teacher might not be rated " poor " in a larger system.

The teachers themselves throw a little light on their difficulties by the reasons they give for them. Table 34 shows how frequently the " good " and " poor " teachers themselves attributed their difficulties to deficiencies in normal school training, lack of experience and other causes. This table is derived from data furnished by the graduates for the other study which has been referred to. The differences between " good " and " poor " teachers are evident in the frequency with which they mention inadequacies of practice teaching, especially the inapplicability of practice teaching to actual conditions. Both groups think lack of experience the most important reason for their difficulties, but " good " teachers place second the lack of instruction in normal school, while " poor " teachers mention second that their practice teaching was inapplicable to actual conditions. If both groups receive the same kind of practice, it is possible that " better " teachers are able to make the necessary " transfer " while " poor " teachers can not. On the other hand, since more of the " poor " teachers were found in rural positions, it is possible that their practice teaching was in fact inadequate to enable them to be as successful as they could have been under other conditions.

TABLE 34

Ratio between number of times various reasons for difficulties are mentioned and number of " good " and " poor " teachers

REASONS GIVEN BY TEACHERS FOR DIFFICULTIES EXPERIENCED	GOOD	POOR
	Ratio to number of teachers	Ratio to number of teachers
Lack of instruction in normal school	.567	.494
Instruction not practical enough	.411	.465
Too little practice teaching	.078	.178
Practice teaching inapplicable to actual conditions	.078	.535
Lack of experience	.700	.717
Miscellaneous	.456	.465
Number of teachers mentioning reasons	90	73

CONCLUSIONS

Low correlations between a criterion and tests designed to predict success may, as has already been stated, indicate either that the criterion is not good or that the tests do not measure the traits that determine success. We should have to conclude from the data at hand that while our criterions are by no means perfect, they apparently measure teaching success better than the entrance tests predict it. Not only in the entrance tests but in all the normal school marks, including practice teaching, we seem to be missing something that is an essential element of successful teaching. There are some indications that this something is closely related to those traits of personality or mental make-up that create what we may designate as *rapport* with children (meaning by this term a composite of sympathy, understanding, insight, ability to gain confidence and thus to control and guide children and hence to avoid situations requiring "discipline"), and that it is much less closely connected with the far more easily measurable traits of scholastic aptitude or achievement.

Why the best teachers have it and the poorest teachers lack it we do not know. Whether it is something innately possessed by some students and not by others, or whether it is a by-product of teacher training which is absorbed by some students and not by others we do not know. We can not even be sure to what extent it may depend on chance conditions of the teaching situation. That success, or at least failure, is judged to a considerable degree by ability to discipline seems clear. How much this ability may depend upon accidental factors of the first few meetings of the teacher with her class, especially in rural situations where no external help is available to bolster up the new teacher while she is finding herself, is a problem that would bear investigation.

It seems probable that teaching success, under differing teaching conditions, may depend upon a variable number of factors which may be combined in *different* ways and still yield success. If this is the case, prognosis of the degree of success would be impossible without a knowledge of the situational elements which would also be determining factors. This point may be made clearer by a symbolic representation. Suppose A and B are the intellectual traits found necessary for teaching success and C and D the personality traits, each of which may be measured and has a maximum value of 10. Suppose R, S, and T are important elements of the teaching situation

which may vary from R_1 to R_n, S_1 to S_n, and T_1 to T_n. Then suppose that $10A + 10B + 10C + 10D + R_1 + S_1 + T_1 = 100$, that is, complete success.

$10A + 10B + 10C + 10D + R_2 + S_2 + T_2$ might yield success only equal to 50, while

$5A + 5B + 5C + 5D + R_3 + S_3 + T_3$ might yield success equal to 75.

If we expand the four traits and the three situational elements to the probable number of elements actually involved, the possibility of the permutations is rather staggering. It suggests, however, that we shall make progress faster if we try to measure success in terms of the teaching situation as well as the qualities of the teacher.

Perhaps the problem of measuring teaching success may be likened to the problem of measuring the success of the driver of a Cadillac and a tractor. This measurement would have to be made by basing it upon the performance of the tractor and the Cadillac in different situations, with different goals and therefore with different criterions of successful performance. We should have to know: (1) what could be expected of each machine; (2) what the aim of any particular performance was; (3) when the aim was accomplished; (4) whether the driver used the machine only for purposes for which it was adapted; (5) how well the machine stood up under his care.

If we carry out the simile, we should measure a teacher's success in terms of: (1) the abilities of the children and the accomplishments possible or reasonable to expect of them; (2) the explicit objective of each unit of teaching; (3) measurement of the accomplishment; (4) the extent to which the teaching met the varying needs, both immediate and remote, of different children; (5) how well the stresses and strains incidental to the child's school activities were anticipated and prevented, or recognized and remedied.

This is of course an oversimplified statement, since the teacher is dealing with organisms, not machines, and is driving from 20 to 40 Rolls Royces, Cadillacs, Fords and tractors all at once instead of one at a time. Nevertheless, it is evident that our school organization does not yet generally permit a teacher to reach all these objectives. Perhaps, therefore, before we solve the problem of measuring teaching success, we shall have to have a better appraisal of the school organization and of the situation into which a teacher is put. In the meantime, probably the most we can do is to continue our effort to

find the things that cause teachers to fail. When these prove to be characteristics of the teacher not easily modifiable by training, we can probably find ways to measure these characteristics and eliminate such persons from training. When they prove to be due to factors in the teaching situation, we should blame neither the teacher nor the training, but seek to modify the conditions of the school organization that defeat the purposes of education.

Bulletins and Pamphlets Prepared by the Educational Research Division

BULLETIN NUMBER

798 **Coxe, W. W.** Silent Reading in New York Rural Schools. 1924. 5 cents

806 **Morrison, J. C., Cornell, W. B. & Coxe, W. W.** Survey of the Need for Special Schools and Classes in Westchester County, New York. 1924. 5 cents

839 **Coxe, W. W. & Cornell, E. L.** A Study of Pupil Achievement and Special Class Needs in Westbury, L. I. 1926. 20 cents

846 **Orleans, J. S. & Richards, E. B.** English in the Rural and Village Schools of New York State. 1926. 10 cents

865 **Orleans, J. S. & Seymour, F. E.** Arithmetic in the Rural and Village Schools of New York State. 1926. 10 cents

875 **Cornell, E. L.** Mental Hygiene: Its Place in the Classroom. 1927. 5 cents

878 **Richards, E. B. & Cornell, E. L.** Suggestions for Teaching English. 1927. 10 cents

892 **Orleans, J. S. & Thompson, H. G.** Survey of Achievement in First Half Year Latin in New York State. 1927. 15 cents

893 **Orleans, J. S.** Manual on the Local Construction and Uses of Objective Tests. 1928. 10 cents

897 **Orleans, J. S. & Thompson, H. G.** Survey of Achievement in Second Half of First Year Latin in New York State. 1928. 10 cents

902 **Orleans, J. S.** Manual of Instruction in Objective Tests for Use at Teachers' Conferences. 1928. 10 cents

906 **Orleans, J. S.** An Experiment in the Local Construction and Uses of Objective Tests in Several School Systems of New York State. 1927. 15 cents

922 **Orleans, J. S. & Baer, J. A.** Handwriting in Rural and Village Schools of New York State. 1929. 20 cents

926 **Coxe, W. W.** Study of the Elementary School Principal in New York State. 1929. 10 cents

927 **Coxe, W. W.** Study of the Secondary School Principal in New York State. 1929. 10 cents

941 **Cowen, P. A.** College Tuition Fees. 1930. 15 cents

944 **Cowen, P. A. & Matthews, E. R.** Special Class Curriculum Study. 1930. 25 cents

948 **Cornell, E. L.** A Beginning Reading Vocabulary for Foreign-Born Adults. 1930. 15 cents

961 **Soper, W. W. & Coxe, W. W.** Trends in Secondary Education. 1930. 20 cents

962 **Soper, W. W. & Baer, J. A.** A Study of the Clerical Duties of District Superintendents of New York State. 1931. 20 cents

964 **Coxe, W. W. & Soper, W. W.** Study of the High School Teacher in New York State. 1931. 15 cents

969 **Coxe, W. W. & Cowen, P. A.** Educational Needs of Pupils in Small High Schools. 1931. 20 cents

975 **Coxe, W. W.** Study of the Elementary School Teacher in New York State. 1931. 5 cents

981 **Cornell, E. L.** The Effect of Trait Differences in Ability Grouping. 1931. 5 cents

BULLETIN
NUMBER

986 Organization of Special Classes for Subnormal Children. 1932. *Revision of Bulletins 802, 819 and 940.* 25 cents

990 **Soper, W. W.** The Use of the Elementary School Auditorium. 1932. 10 cents

1001 **Coxe, W. W.** Levels and Ranges of Ability in New York State High Schools. 1932. 5 cents

1007 **Cowen, P. A.** College Entrance Inquiry. 1932. 40 cents

1019 **Soper, W. W.** Development of State Support of Education in New York State. 1933. 25 cents

1025 **Coxe, W. W.** Directed Study. 1933. 10 cents

1029 **Soper, W. W. & Hollister, F. J.** The Postgraduate Problem in New York State High Schools. 1934. 10 cents

1033 **Coxe, W. W. & Cornell, E. L.** The Prognosis of Teaching Ability of Students in New York State Normal Schools. 1934. 25 cents

---

Bibliography of Objective Tests in the Vocational Subjects. 1933. 5 cents

Bibliography of Research Studies in Education Pertaining to New York State. 1932. 10 cents

Needed Educational Research in New York State and Research Studies in Progress, 1931–32. Prepared under the direction of the Committee on Cooperation in Educational Research. 1932. 10 cents

Needed Educational Research in New York State. Prepared under the direction of the Committee on Cooperation in Educational Research. 1933. 10 cents

---

Publications of the State Education Department are free to schools and libraries in New York State. Others may purchase limited quantities at the cost of publication listed above.

The above bulletins and pamphlets should be ordered from and checks made payable to the New York State Education Department, Albany, N. Y.

3 0112 105869777